Discrete Applied Mathematics 26 (1990) 137-157
North-Holland

TWO MIXED INTEGER PROGRAMMING FORMULATIONS ARISING IN MANUFACTURING MANAGEMENT*

Robert G. JEROSLOW

Received 5 June 1987
Revised 17 May 1988

We provide formulation techniques for obtaining sharp (i.e., convex hull) mixed integer programming (MIP) formulations for different classes of lot-sizing problems, including various backlogging models, and with the possibility of fixed charges on the production, backlogging, and inventory variables.

We also explore approximation results for MIP formulations of generalized shop loading problems.

Keywords. Mixed integer programming, disjunctive methods, variable redefinition, lot sizing, shop loading.

1. Introduction

In recent years, there has been a growing awareness of the importance of model formulation and knowledge representation in mixed integer programming (MIP). While the theoretical frameworks for algorithms have become fairly standardized, including branch-and-bound techniques, Lagrangean relaxation approaches, decomposition methods and cutting-plane methods, theoretical work in MIP problem formulation is more recent. New discoveries continue to emerge which have significant practical importance.

One of the active areas of research, which finds application to problems occurring in production and in operations management, concerns networks with fixed charges on arcs (e.g., [5-7, 12-14, 25, 27, 36-38, 40, 47]). The dynamic lot-sizing problem [44, 45, 48, 49] is one which can be cast in this framework, and advances have been obtained with dramatic consequences for computation [14, 37].

I have been working, jointly and individually, on a general approach to MIP representability [18-22] derived from Meyers' ideas [31-33], Ibaraki [16], and the disjunctive methods [2-4, 9, 10, 17]. From this perspective, it is desirable to include the seemingly ad hoc reformulation devices for lot-sizing problems in such a frame-

* This research has been partially supported by NSF grant DMS-8513970. This paper has been prepared in connection with the conference Combinatorial Optimization 1987, held at the University of Southampton, April 4-6, 1987.

work. Martin's "variable redefinition" method represents a very novel departure with potential for a broad range of applications. In [21, Lecture 4] we provided a result which generalized Martin's variable redefinition method together with our earlier methods.

In this paper, we present a stronger result (Theorem 3.1) and show that it can be used to obtain a sharp (tight) formulation for one version of lot sizing extended to allow backlogging, as explored recently in [37]. As demonstrated in an example below, our approach also allows other treatments of backlogging, and other kinds of fixed charges than only those associated with production activities.

In Section 2, we summarize the requisite background from our general approach to MIP representability. In Section 3, we establish a broad principle (Theorem 3.1) which unifies this approach with "variable redefinition", and then we proceed to apply it in the setting of networks. We also provide a corollary which is a different kind of application (Corollary 3.4).

As it develops, the broad principle we cite has to do with compact, sharp formulations of unions of sums of sets, in which the union may be over an exponential (or even infinite) index set. In Section 4, we complement our main results with a result (Theorem 4.2) on sums of unions of sets. It is an approximate result in the spirit of the Shapley–Folkman–Starr theorem [1] and similar ideas have been used elsewhere (e.g., [8, 11, 27]). It also has application in the production area.

We conclude this section with some comments on mathematical notation.

For a set $S \subseteq R^p$, we used conv(S) respectively cl conv(S) to denote the convex span respectively the closure of the convex span of S. For other notation and results on convexity we use [39] or [42] as general references.

2. Preliminaries

A set $S \subseteq R^p$ is called *b-MIP.r* (for "bounded-MIP-representable") if there is a linear transformation $g(x; u) = Ax + Bu$ (compactly represented as $g(x; u)$), possibly in auxiliary variables u, a vector b, and an index set K (possibly $K = \emptyset$) of the set of indices of u, such that:

$$x \in S \quad \Leftrightarrow \quad \begin{array}{l} \text{there exists } u \text{ with } u_k \in \{0,1\} \text{ for } k \in K, \text{ and} \\ g(x; u) \le b. \end{array} \tag{2.1}$$

The right-hand side of (2.1) is called a *representation* of S.

The *relaxation* of the representation of $\underline{S}$ in (2.1) of S, is defined by:

$$\mathrm{Rel}(\underline{S}) = \{x \mid \text{for some } u \text{ with } 0 \le u_k \le 1 \text{ for } k \in K,\ g(x; u) \le b\}. \tag{2.2}$$

Generally, Rel($\underline{S}$) depends on the representation $\underline{S}$ for S. (Typically, the linear inequalities "$0 \le u_k \le 1$ for $k \in K$" are taken to be already included in $g(x; u) \le b$.)

The *starred recession cone* rec*($\underline{S}$) is defined by:

$$\mathrm{rec}^*(\underline{S}) = \{x \mid \text{for some } u \text{ with } u_k = 0 \text{ for } k \in K,\ g(x; u) \le 0\}. \tag{2.3}$$

Clearly $\operatorname{rec}^*(\underset{\sim}{S})$ is a polyhedral cone; in particular, $\alpha \operatorname{rec}^*(\underset{\sim}{S}) = \operatorname{res}^*(S)$ for $\alpha > 0$.

For future reference, we recall this definition of $\operatorname{rec}(S)$ from [39, 42]:

$$\operatorname{rec}(S) = \{x \mid \text{for some } x' \in S \text{ and all } \lambda \geq 0,\ x' + \lambda x \in S\}. \tag{2.4}$$

We also recall that, if $S \neq \emptyset$ is closed and convex:

$$\operatorname{rec}(S) = \{x \mid \text{for all } x' \in S \text{ and all } \lambda \geq 0,\ x' + \lambda x \in S\}. \tag{2.5}$$

Theorem 2.1 [20]. *If $S \neq \emptyset$, then*

$$\operatorname{rec}^*(\underset{\sim}{S}) = \operatorname{rec}(S). \tag{2.6}$$

In particular, $\operatorname{rec}^*(\underset{\sim}{S})$ depends only on S, and not on $\underset{\sim}{S}$, if $S \neq \emptyset$. Let $\mathrm{RL}(\underset{\sim}{S})$ denote the representation for $\operatorname{Rel}(\underset{\sim}{S})$ as in the bracketed definition in (2.2). By combining (2.2) and (2.3), one easily proves that:

$$\operatorname{rec}^*(\mathrm{RL}(\underset{\sim}{S})) = \operatorname{rec}^*(\underset{\sim}{S}). \tag{2.7}$$

Thus $\operatorname{rec}^*(\mathrm{RL}(\underset{\sim}{S})) = \operatorname{rec}(S)$, if $S \neq \emptyset$.

The following result characterizes b-MIP.r sets S.

Theorem 2.2 [22]. *A set S is b-MIP.r exactly if S is a finite union of polyhedra $S = P_1 \cup \cdots \cup P_t$ with the same recession cone (i.e., $\operatorname{rec}(P_i)$ is independent of i, $1 \leq i \leq t$).*

In [20] the equivalence of (2.4) and (2.5) was shown for any b-MIP.r set S.

A representation (2.1) is called *sharp* if

$$\operatorname{Rel}(\underset{\sim}{S}) = \operatorname{conv}(S). \tag{2.8}$$

We always have $\operatorname{Rel}(\underset{\sim}{S}) \supseteq \operatorname{conv}(S)$. In [22] we showed that every b-MIP.r set has a sharp representation. In this paper, we focus on the *size* of sharp representations.

3. Results on unions of sums

For the main result of this section (Theorem 3.1), we will need the following notation.

A set $S \subseteq R^p$ is given with a representation

$$x \in S \quad \Leftrightarrow \quad \begin{array}{l} \text{there exists } u \text{ with } u_k \in \{0,1\} \text{ for } k \in K, \text{ and} \\ g(x; u) \leq b. \end{array} \tag{3.1}$$

Similarly, sets $T_i \subseteq R^q$ are given for $i = 1, \ldots, p$ with representations:

$$v \in T_i \quad \Leftrightarrow \quad \begin{array}{l} \text{there exists } w^{(i)} \text{ with } w_k^{(i)} \in \{0,1\} \text{ for } k \in K_i, \text{ and} \\ f_i(v; w^{(i)}) \leq d^{(i)}. \end{array} \tag*{(3.2)$_i$}$$

We define the set $V\subseteq R^q$ by:

$$V=\bigcup_{x\in S}\left\{\sum_{i=1}^{p} x_i t^{(i)}\,\middle|\, t^{(i)}\in T_i \text{ for } i=1,\dots,p\right\},\tag{3.3}$$

i.e., $V=\bigcup_{x\in S}\sum_i x_i T_i$. In (3.3), we observe the *nonstandard convention* that $0.T_1+\cdots+0.T_p=\operatorname{rec}(T_1)$ (needed to treat the case $0\in S$; see below).

Our main result states sufficient conditions under which a possibly very large (and possibly even infinite) union of representable sets has a relatively succint *sharp* representation. An earlier (but weaker) result of this type was announced in [21].

Theorem 3.1. *Suppose that these hypotheses hold:*

(1) $S\subseteq R^p_+$, $S\neq\emptyset$;
(2) *each* $T_i\neq\emptyset$, *and* $\operatorname{rec}(T_i)$ *is independent of* $i=1,\dots,p$;
(3) *whenever* $K_i\neq\emptyset$, *then* $x_i\in\{0,1\}$ *for all* $x\in S$;
(4) *the representations* (3.1) *and* $(3.2)_i$ *for* $i=1,\dots,p$ *are sharp.*

Then the following is a sharp representation of V:

$$\begin{aligned} v\in V \;\Leftrightarrow\;\; & \textit{there are } u, w^{(1)},\dots,w^{(p)},x,v^{(1)},\dots,v^{(p)} \textit{ with } u_k\in \\ & \{0,1\} \textit{ for } k\in K,\; w_k^{(i)}\in\{0,1\} \textit{ for } k\in K_i,\; i=1,\dots,p, \\ & \textit{and } g(x;u)\le b,\; f_i(v^{(i)};w^{(i)})\le d^{(i)}x_i \textit{ for } i=1,\dots,p, \\ & 0\le w_k^{(i)}\le x_i \textit{ for } k\in K_i \textit{ and } i=1,\dots,p,\; v=\textstyle\sum_{i=1}^p v^{(i)}. \end{aligned}\tag{3.4}$$

Proof. First, we show that (3.4) is a representation of V (this part does not require the hypotheses (4) on sharpness).

Clearly, if $v\in V$, then either: (a) $v=\sum_{i=1}^p x_i t^{(i)}$ with $x\in S$, $x\neq 0$, and $t^{(i)}\in T_i$ for $i=1,\dots,p$; or (b) $0\in S$ and $v\in\operatorname{rec}(T_1)$.

For the case (a), note that u exists satisfying (3.1) and that $w_0^{(i)}$ exists satisfying (3.2) with $v=t^{(i)}$, whenever $x_i>0$. For $x_i=0$, let $w_*^{(i)}$ be such that $f_i(0,w_*^{(i)})\le 0$ and $w_{*_k}^{(i)}=0$ for $k\in K_i$. Since $0\in\operatorname{rec}^*(T_i)$, $w_*^{(i)}$ exists. (Note that $x_i<0$ is impossible since $S\subseteq R^p_+$ by hypothesis.)

Put $v^{(i)}=x_i t^{(i)}$ for $x_i>0$ and $v^{(i)}=0$ for $x_i=0$. Put $w^{(i)}=x_i w_0^{(i)}$ for $x_i>0$ and $w^{(i)}=w_*^{(i)}$ for $x_i=0$. From $f_i(t^{(i)};w_0^{(i)})\le d^{(i)}$ and homogeneity, we have $f_i(v^{(i)};w^{(i)})\le d^{(i)}x_i$ if $x_i>0$. Moreover, for $x_i=0$ we have $f_i(v^{(i)};w^{(i)})=f_i(0;w_*^{(i)})\le 0=d^{(i)}x_i$ by construction. If $x_i>0$ and $K_i\neq\emptyset$, then by hypothesis $x_i=1$, and hence $w_k^{(i)}\le x_i$ for $k\in K_i$ by construction. Thus $0\le w_k^{(i)}\le x_i$ for $k\in K_i$ and $i=1,\dots,p$. Moreover,

$$v=\sum_{i=1}^p x_i t^{(i)}=\sum_i\{v^{(i)}\mid x_i>0\}+\sum_i\{v^{(i)}\mid x_i=0\}.$$

Thus, the r.h.s. of (3.4) holds.

For case (b), put $x=0$, $v^{(1)}=v$, $v^{(i)}=0$ for $i\ge 2$; and let $w^{(i)}$ be such that $f_i(v^{(i)};w^{(i)})\le 0$ with $w_k^{(i)}=0$ for $k\in K_i$. Again, u exists with $g(0;u)\le b$ and $u_k\in\{0,1\}$ for $k\in K$. Clearly $w_k^{(i)}\le x_i$ for $k\in K_i$ and $i=1,\dots,p$. Again, the r.h.s. of (3.4) holds.

For the converse, suppose the r.h.s. of (3.4) holds. Then either: (c) $x \neq 0$; or (d) $x=0$.

For case (c), assume without loss of generality that $x_1 > 0$. If $x_i > 0$, $f_i(v^{(i)}/x_i;\ w^{(i)}/x_i) \leq d^{(i)}$ by homogeneity, and hence $t^{(i)} = v^{(i)}/x_i \in T_i$ (note that $x_i = 1$ if $K_i \neq \emptyset$, so $w_i^{(k)}/x_i = w_i^{(k)} \in \{0,1\}$ for $k \in K_i$). If $x_i = 0$, then $w_k^{(i)} = 0$ for $k \in K_i$, so $v^{(i)} \in \mathrm{rec}(T_i)$ $(=\mathrm{rec}(T_1))$, since $f_i(v^{(i)};\ w^{(i)}) \leq 0$. Thus $v' = \sum_{x_i=0} v^{(i)} \in \mathrm{rec}(T_1)$ and also $v = \sum_{x_i>0} x_i t^{(i)} + v'$. We have $v = x_1(t_1 + v'/x_1) + \sum_{i \neq 1,\ x_i > 0} x_i t^{(i)} \in V$ since $t_1 + v'/x_1 \in T_1$ (as $v'/x_1 \in \mathrm{rec}(T_1)$).

For case (d), clearly $0 \in S$ and also $w_k^{(i)} = 0$ for $k \in K_i$ and $i = 1, \dots, p$. Then as $f_i(v^{(i)};\ w^{(i)}) \leq 0$, $v^{(i)} \in \mathrm{rec}(T_1)$ for $i = 1, \dots, p$ so $v = \sum_i v^{(i)} \in \mathrm{rec}(T_1) \subseteq V$ (using the nonstandard convention regarding $0.T_1 + \cdots + 0.T_p$).

We have thus established that (3.4) represents V. We now show that (3.4) is a sharp representation. Toward that end, it suffices to show that, in the relaxation of the r.h.s. of (3.4), we necessarily have $v \in \mathrm{conv}(V)$.

As in the proof (just completed) of the converse implication, we will have (by hypothesis (4)) $x \in \mathrm{conv}(S)$ and also $t^{(i)} \in \mathrm{conv}(T_i)$ when $x_i > 0$, while $v^{(i)} \in \mathrm{rec}(T_i)$ if $x_i = 0$. Write $x = \sum_k \lambda_k x^{(k)}$ with each $x^{(k)} \in S$, all $\lambda_k \geq 0$, $\sum_k \lambda_k = 1$. Then $v = \sum_{x_i>0} x_i t^{(i)} + v'$ where $v' = \sum_{x_i=0} v^{(i)} \in \mathrm{rec}(T_1)$ by hypotheses. We have

$$
\begin{aligned}
v &= \sum_{x_i>0} \Big(\sum_k \lambda_k x_i^{(k)} \Big) t^{(i)} + v' = \sum_k \lambda_k \Big(\sum_{x_i>0} x_i^{(k)} t^{(i)} \Big) + v' \\
&= \sum_k \lambda_k \Big(\sum_{x_i>0} x_i^{(k)} t^{(i)} + v' \Big).
\end{aligned}
$$

If $x^{(k)} \neq 0$, we show that $\sum_{x_i>0} x_i^{(k)} t^{(i)} + v' \in \mathrm{conv}(v)$ as in case (c). If $x^{(k)} = 0$, then we show that $\sum_{x_i>0} x_i^{(k)} t^{(i)} + v' = v' \in \mathrm{conv}(V)$ as in case (d). Thus $v \in \mathrm{conv}(V)$, as desired. □

In the applications found for Theorem 3.1 to date, typically the sets T_i are polyhedral or (as in variable redefinition) even singleton sets, for which the sharpness hypotheses and recession hypotheses of Theorem 3.1 are trivially verified. From this perspective, the content of Theorem 3.1 is that the crucial hypothesis is sharpness of the given representation for the set S. Typically, S is the set of incidence vectors of a "hidden index set", as e.g. the incidence vectors of the set of paths in a graph. The presence of such an indexing set is usually not at all obvious from standard representations for V.

Corollary 3.2. *With the hypotheses of Theorem* 3.1, *define* T_i' *by*

$$
T_i' = \{ v \mid f_i(v;\ w^{(i)}) \leq d^{(i)},\ w_k^{(i)} = 0 \ \textit{for}\ k \in K_i \}. \tag{3.5}
$$

Then if each $T_i' \neq \emptyset$,

$$
\mathrm{rec}(T_1) = \mathrm{rec}(T_i'), \tag{3.6}
$$

and moreover

$$\operatorname{rec}(V)=\bigcup_{x\in\operatorname{rec}(S)}\left(\sum_i x_i T_i'\right). \tag{3.7}$$

Proof. Note that $T_i'\subseteq T_i$ is a polyhedron with a representation (3.5) having no binary-constrained variables, so that hypothesis (3) of Theorem 3.1 trivially holds for the T_i' in place of the T_i. Hypothesis (1) holds for rec(S) in place of S, since $S\subseteq R_+^p$ and $0\in\operatorname{rec}(S)$. We verify (3.6) using (2.3), (3.4) and (3.5), so that $\operatorname{rec}(T_i')$ is independent of i. Since all $T_i'\neq\emptyset$, this verifies hypothesis (2) for the T_i' in place of the T_i. Finally, polyhedral representations are always sharp, which verifies hypothesis (4) for rec(S) and the T_i' in place of S and the T_i.

We have just verified the hypothesis of Theorem 3.1 for rec(S) and the T_i' in place of S and the T_i. Therefore, by Theorem 3.1, if the r.h.s. of (3.4) has b replaced by the zero vector and the constraints "$u_k=0$ for $k\in K$, $w_k^{(i)}=0$ for $k\in K_i$ and $i=1,\dots,p$" are appended, this r.h.s. describes the r.h.s. of (3.7). However, by (2.3) this r.h.s. also represents $\operatorname{rec}^*(\underline{V})=\operatorname{rec}(V)$. □

We remark that the additional hypothesis in Corollary 3.2, that all $T_i'\neq\emptyset$, is not restrictive. Given that $T_i'\neq\emptyset$, by complementing binary variables, as necessary (possibly changing $d^{(i)}$), it will always hold.

See [19] for a corollary of Theorem 3.1 which is of a different nature than the lot-sizing applications we emphasize here.

We next turn to applications of Theorem 3.1 in lot-sizing problems.

Suppose now that a linear system is given which describes a polytope (i.e., bounded polyhedron)

$$f(y)\leq b,\quad y\geq 0, \tag{3.8}$$

and that it is desired to add binary variables z_q for each y_q with the property that

$$y_q>0 \quad\text{implies}\quad z_q=1. \tag{3.9}$$

For example, (3.9) is often required in order to assess a "fixed charge" $f_q\geq 0$ for having $y_q\geq 0$ (i.e., "$+f_q z_q$" will be added to the cost-minimizing criterion function).

Our next result provides a basic principle for helping to obtain a sharp description of the set T of points (y,z) defined by (3.8) and (3.9) (including the requirement of binary z). It states that the relaxation of a sharp description of the (y,z) vectors associated with the extreme points of (3.8) is conv(T), and that to obtain a representation of T one need only add the binary requirement on z to this relaxation. This principle will be sufficient for our needs in this paper, although it can be extended by consideration of general extreme subsets of (3.8). The principle reduces our work to that of finding a compact, sharp description of T restricted to extreme points of (3.8).

Lemma 3.3. *Suppose that* (3.8) *describes a bounded polyhedron with extreme points* $y^{(1)}, \dots, y^{(s)}$. *Define the points* $z^{(1)}, \dots, z^{(s)}$ *and the sets* $W_1, \dots, W_s$ *by the conditions*:

$$z_q^{(j)} \in \{0,1\} \text{ and } z_q^{(j)} = 1 \text{ exactly if } y_q^{(j)} > 0,\ j = 1, \dots, s, \tag{3.10}$$

$$W_j = \{(y,z) \mid y = y^{(j)},\ z \geq z^{(j)},\ z \text{ binary}\}. \tag{3.11}$$

Finally, let a sharp representation of $W = \bigcup_j W^{(j)}$ *be given by*

$$h(y,z;u) \leq d,\ u_k \in \{0,1\} \text{ for } k \in K. \tag{3.12}$$

Then a sharp representation of the set T *defined by* (3.8) *and* (3.9) *is*

$$h(y,z;u) \leq d,\ 0 \leq u_k \leq 1 \text{ for } k \in K,\ z \text{ binary}. \tag{3.13}$$

Proof. The linear relaxations of (3.13) and (3.12) are the same. Hence an extreme point of the projection onto (y,z) of the linear relaxation of (3.13) lies in some $W^{(j)}$, and $W^{(j)} \subseteq T$. Thus if we prove that (3.13) represents T, we are done.

If $(y,z) \in T$, then for some $\lambda_j \geq 0$, $\sum_j \lambda_j = 1$, we have $y = \sum_j \lambda_j y^{(j)}$. For each $j = 1, \dots, s$ if $\lambda_j > 0$, then $(y^{(j)}, z) \in W_j$. In fact, if $\lambda_j > 0$ and $y_q^{(j)} > 0$, then also $y_q > 0$, so $z_q = 1$ and thus $z \geq z^{(j)}$.

We see that $(y,z) = \sum_{\lambda_j > 0} \lambda_j (y^{(j)}, z) \in \operatorname{conv}(W)$, and hence there is u with $0 \leq u_k \leq 1$, for $k \in K$ and $h(y,z,u) \leq d$. Since also $z_q \in \{0,1\}$ for all g, (3.13) is satisfied.

Conversely, if (3.13) is satisfied, by the sharpness hypothesis on (3.12) we have $(y,z) \in \operatorname{conv}(W)$. Thus $y \in \operatorname{conv}(\{y^{(1)}, \dots, y^{(s)}\})$ so $f(y) \leq b$. Also, since $(y,z) \in \operatorname{conv}(W)$, if $y_q > 0$, then $z_q > 0$. However, as $z_q \in \{0,1\}$ also, (3.12) holds and $(y,z) \in T$. $\square$

Since the number of extreme points of (3.8) often can be very large, we are not a priori guaranteed of a compact description of W in (3.12). In order to obtain a compact description, several other structural features can be useful, which will allow us to use Theorem 3.1. In some directed network problems, for example, the set W of Lemma 3.3 can be constructed as having the additive form $\sum_i x_i T_i$ of (3.3), for a suitable set S, as we shall see in the examples below.

Our next result is well known (e.g. [49]). We provide a proof of it for the sake of completeness. For backlogging treated as negative inventory, Zangwill [49] used this result as a basis for a dynamic programming algorithm. This is also the version of backlogging treated by Pochet and Wolsey [37]. For network flows, (3.8) consists of node equations for flow conservation. However, since total demand is known, every arc can be viewed as having (an implicit) capacity bound.

Lemma 3.4. *Suppose that* (3.8) *describes a directed, single-commodity network in which certain variables among y are sources, and all constraints are either flow conservations at nodes or outflows of given amounts at sinks.* (*In particular, while inflow variables occur at sources, their values are not specified.*)

Then in an extreme point $\bar{y}$ *of* (3.18), *positive variables* $y_q > 0$ *occur on a set of* *arcs which forms a union of node-disjoint directed trees, with each tree rooted at a unique source node with positive flow, having a set of sink nodes as leaves and with any sink node occurring in some tree.*

Proof. We claim that no *undirected* loop can occur in the positive arcs of an extreme point solution. In fact, if such a loop were to occur, an orientation can be chosen as can a positive amount $\delta>0$ ($\delta=$ minimum flow in any arc of the loop), such that if all positive arcs (i.e., arcs with the given orientation) have flow increased by δ, and all negative arcs have flow decreased by δ, as feasible flow results. Moreover, a feasible flow results by decreasing positive flows by δ and increasing negative flows by δ. However, the given flow is the average of the two distinct feasible flows just described, so it would be an extreme solution.

We also claim that, in the positive arcs of an extreme point solution, one cannot find an *undirected* path between two source nodes which both have positive flow. The argument for this claim is similar to that of the last paragraph, involving a choice of orientation of the supposed path, and increases and decreases of source variables as well as arc variables; we leave the details to the reader.

From the two claims, the positive variables of an extreme point solution do form a union of undirected node-disjoint trees, with at most one source node having positive flow in each tree. Moreover, each tree is also a directed tree. Indeed, if two incoming arcs to a node existed, as both derive flow from the same source node, an undirected loop would occur, which cannot be the case. □

We conclude this section with two examples. In each case, we first describe the lot-sizing situation, and the kind of network involved in each. We then apply Lemma 3.4 to obtain a description of the extreme points of the set of feasible flows, and use Lemma 3.3 as the principle by which we shall obtain a sharp description of the flows *plus* fixed charges. In order to make this sharp description compact, Theorem 3.1 will be used.

Example 3.5. In our first example, we shall augment the simple lot-sizing model of [45] by backlogging, and we shall treat backlogging as negative inventory (with a generally different unit cost and fixed charge for negative as opposed to positive inventory).

In this setting, a single product is manufactured to meet known positive demands in each of r time periods. Excess manufacture may be inventoried into later periods. By the end of the planning horizon, all demands must be satisfied. However, in a given period demand may fail to be met from previous and current manufacture, and may be backlogged to be filled by manufacture in later periods.

The activities (variables) of this model are manufactured in period j, denoted u_j; inventory carried forward from period j to $(j+1)$, denoted I_j; demand backlogged from period $(j+1)$ to period j (which may in turn be backlogged again), denoted s_j. We also use variables $w_j=d_j$ ($d_j=$ demand in period j) which are not essential in this setting, but which will have a use in Example 3.6 below. All variables incur both per unit flow costs and fixed charges. The network involved for r periods is depicted in Fig. 1(a) and the equations (flow conservation at nodes) for this network are as follows (with all variables nonnegative):

$$\begin{aligned} &u_1+s_1=w_1+I_1, \\ &u_j+I_{j-1}+s_j=w_j+I_j+s_{j-1}, \quad j=2,\dots,r-1, \\ &u_r+I_{r-1}=w_r+s_{r-1}. \end{aligned} \tag{3.14}$$

The variables in (3.14) form the vector y of (3.8) to which fixed charge variables are added as in (3.9).

The directed trees mentioned in Lemma 3.4 are as follows in Fig. 1. There will be a pair of indices $1 \le j \le j' \le r$ (possibly $j=j'$) and an index $j \le i \le j'$ such that all demands for periods j through j' are met by production in period i. Demands for periods $j, j+1, \dots, i-1$ are met by backlogging from period i; demand for period i is met directly from manufacture in period i; and demands for periods $i+1, \dots, j'$ are met by inventory carried forward from period i. In such a tree, flows $s_j, s_{j+1}, \dots, s_{i-1}, u_i, I_i, I_{i+1}, \dots, I_{j'-1}$ are all positive as are $w_j, \dots, w_{j'}$; all other arcs entering or leaving the tree have zero flow. Unique values for the positive flows can easily be computed by starting at the leaves of the tree and working toward its root. Moreover, any demand arc is part of exactly one such directed tree, so that extreme point solutions are a union of such trees, as described in Lemma 3.4.

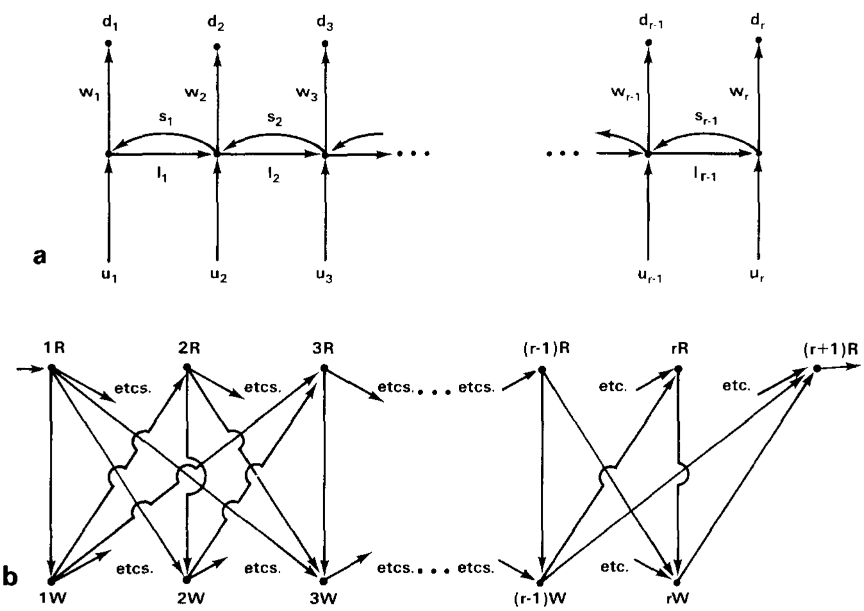

Fig. 1.

We view that an extreme point solution is determined by a series of decisions, beginning with the choice of the first tree $j_1 = 1 \le i_1 \le j_1'$, followed by the choice of the second tree $j_2 = j_1' + 1 \le i_2 \le j_2'$, the choice of the third tree $j_3 = j_2' + 1 \le i_3 \le j_3', \dots$, etc., on to the choice of the last tree $j_t = j_{t-1}' + 1 \le i_t \le j_t' = r$. However, for each tree with $j_k < i_k$ it is expedient to view the choice $j_k \le i_k \le j_k'$ as taken in two steps: First a choice of j_k and i_k, followed by a choice of j_k'. This latter conceptualization will allow us a reduction in the number of auxiliary variables to quadratic order in r, as opposed to cubic order in r.

The sequence of decisions involved in an extreme point solution can be formalized as a path in a suitable graph, which we now describe. To each time period $t \in \{1, \dots, r\}$ there will be two nodes t(R) and t(W) ("red" and "white"). We shall also need a node $(r+1)$(R). An arc from j(R) to i(W), $j \le i$, will symbolize the choice to backlog demand from periods $j, j+1, \dots, i-1$ from manufacture in period i. The arcs from j(R) to j(W) have only a formal role to allow return to "white" nodes, and are used when the next actual choice does not involve backlogging. An arc from i(W) to $(j'+1)$(R), $i \le j'$, will symbolize the choice to meet demand in periods $i, i+1, \dots, j'$ from manufacture in period i, thus incurring inventory carried into periods $i+1, \dots, j'$. These are the only kinds of arcs which occur in the network. Clearly, a path through this network from sole source node 1(R) to sole sink node $(r+1)$(R) corresponds to an extreme point solution. See Fig. 1(b).

To formalize this construction in a polyhedral manner, a binary variable $x_{(g,h)} = x(g,h)$ is associated with every arc (g,h) of the network described, and the set S of all incidence vectors of paths from source to sink is the "hidden index set" we mentioned previously. In order to apply Theorem 3.1, we shall associate a b-MIP.r $T_{(g,h)} \overset{\text{def}}{=} T(g,h)$ with each arc (g,h), in a manner we describe next. This polytope describes that portion of a tree associated with the arc (g,h).

All $T(g,h)$ shall contain vectors with components for all of the flow variables u_j, w_j, s_j, I_j for all $j = 1, \dots, r$, as well as for their associated fixed charge variables $z(u_j)$, $z(w_j)$, $z(s_j)$, $z(I_j)$ for $j = 1, \dots, r$.

If $g = j$(R) and $h = i$(W), the equations describing $T(g,h)$ are:

$$
\begin{array}{lll}
u_k = 0, & z(u_k) = 0, & \text{for } k = 1, \dots, j-1, i+1, \dots, r, \\
u_k = 0, & 1 \ge z(u_k) \ge 0, & \text{for } k = j, \dots, i-1, \\
u_i = \sum_{j \le p \le i-1} d_p, & z(u_i) = 0, & \\
w_k = 0, & z(w_k) = 0, & \text{for } k = 1, \dots, j-1, i, \dots, r, \\
w_k = d_k, & z(w_k) = 1, & \text{for } k = j, \dots, i-1, \\
s_k = 0, & z(s_k) = 0, & \text{for } k = 1, \dots, j-1, i, \dots, r, \\
s_k = \sum_{j \le p \le k} d_p, & z(s_k) = 1, & \text{for } k = j, \dots, i-1, \\
I_k = 0, & 1 \ge z(I_k) = 0, & \text{for } k = j, \dots, i-1,
\end{array}
\tag{3.15}
$$

$$I_k=0, \qquad z(I_k)=0, \quad \text{for } k=1,\dots,j-1,i,\dots,r,$$

all z variables binary. (Note that (3.15) is the zero vector when $j=i$.)

If $g=i(\mathrm{W})$ and $h=(j'+1)(\mathrm{R})$, the equations describing $T(g,h)$ are:

$$\begin{aligned}
&u_k=0, && z(u_k)=0, && \text{for } k=1,\dots,i-1,j'+1,\dots,r,\\
&u_k=0, && 1\geq z(u_k)\geq 0, && \text{for } k=i+1,\dots,j',\\
&u_i=\sum_{i\leq p\leq j'} d_p, && z(u_i)=1, && \\
&w_k=0, && z(w_k)=0, && \text{for } k=1,\dots,i-1,j'+1,\dots,r,\\
&w_k=d_k, && z(w_k)=1, && \text{for } k=i,\dots,j',\\
&s_k=0, && z(s_k)=0, && \text{for } k=1,\dots,i-1,j'+1,\dots,r,\\
&s_k=0, && 1\geq z(s_k)\geq 0, && \text{for } k=i,\dots,j',\\
&I_k=0, && z(I_k)=0, && \text{for } k=1,\dots,i-1,j',\dots,r-1,\\
&I_k=\sum_{k<p\leq j'} d_p, && z(I_k)=1, && \text{for } k=1,\dots,j'-1,
\end{aligned} \tag{3.16}$$

all z variables binary.

As we next see, the liberal use of zero settings for variables not associated with the decision that the arc (g,h) represents, allows the actual value of those variables to be filled in by addition along the arc associated with a decision affecting those variables.

The sets $T(g,h)$ are all sharp descriptions and have zero as recession cone. Moreover, $T(j(\mathrm{R}),i(\mathrm{W}))+T(i(\mathrm{W});j'(\mathrm{R}))$ is a representation of the tree corresponding to a choice of production in period i to supply periods j through j', $j\leq i\leq j'$. Finally, if x is the indicator vector of a path from source to sink, $\sum x(g,h)T(g,h)$ is $W^{(j)}$ in the sense of Lemma 3.3. Thus, a compact sharp description (3.12) of $W=\bigcup_j W^{(j)}$ is one for $V=\bigcup_{x\in S}(\sum_i x_i T_i)$ of (3.3), and we may use the representation in (3.4) of Lemma 3.1.

Significant simplifications occur in (3.4) due to the fact that all inequalities in (3.15) and (3.16) are equations, with the exception of those inequalities for fixed charge variables which potentially could have been set to unity by some decision leaving from node g, but were not along arc (g,h); these latter variables alone are permitted to be zero or one. Allowing for the fact that even these latter variables have a setting (of zero) at an optimal solution for nonnegative fixed charges, one can entirely substitute out for $T(g,h)$ in (3.4), so that the inequalities $f_i(v^{(i)};\, w^{(i)})\leq d^{(i)}x_i$ for $i=(g,h)$ in (3.4) do not explicitly occur. Moreover, by use of these zero settings, the auxiliary variables $w_k^{(i)}$ of T_i, $i=(g,h)$, are set to zero, and so the inequalities $0\leq w_k^{(i)}\leq x_i$ in (3.4) also need not occur. Finally, the equation $v=\sum_{i=1}^{p} v_i$ of (3.4) can be used to substitute out for all of the variables v (i.e., u_j, s_j, w_j, I_j, $z(u_j)$, $z(s_j)$, $z(w_j)$, $z(I_j)$) in terms of the variables $v^{(i)}=d^{(i)}x_i$, $i=(g,h)$, where $d^{(i)}$ is the singleton set described in either (3.15) or (3.16) for $i=(g,h)$ with binary variables set to zero which are not otherwise set.

The resulting expressions for variables v may be used where those variables occur in the objective function and (except for fixed charge variables) in any additional constraints of the MIP other than those from the lot-sizing constraints. Thus, out of (3.4), only the constraints $g(x,u)\le b$ remain. As these constraints are simply those for a sharp formulation of the incidence vectors of paths from source to sink, these can be obtained by putting a unit flow into the source and requiring a unit flow out of the sink, and writing the node conservation equations. The unimodularity property of network flows then assures sharpness [15]. In particular, no auxiliary variables u are needed, and also explicit unit upper bounds on arc flows are not necessary due to the unit total inflow.

The resulting inequalities $g(x;u)\le b$ thus involve $2r+1$ node conservation equations in the $r(r+1)$ variables $x(g,h)$. A sharp formulation with $r+1$ constraints but $O(r^3)$ auxiliary variables can also be obtained; we omit details.

For example, with $r=3$ time periods, using a unit flow in and out of the network of Fig. 1(b), we see that the constraints $g(x;u)\le b$ are nonnegativities plus the node conservation equations, as follows:

$$
\begin{aligned}
&x(1\mathrm{R},1\mathrm{W})+x(1\mathrm{R},2\mathrm{W})+x(1\mathrm{R},3\mathrm{W})=1 \quad (\text{node } 1\mathrm{R}),\\
&-x(1\mathrm{R},1\mathrm{W})+x(1\mathrm{W},2\mathrm{R})+x(1\mathrm{W},3\mathrm{R})+x(1\mathrm{W},4\mathrm{R})=0 \quad (\text{node } 1\mathrm{W}),\\
&-x(1\mathrm{W},2\mathrm{R})+x(2\mathrm{R},2\mathrm{W})+x(2\mathrm{R},3\mathrm{W})=0 \quad (\text{node } 2\mathrm{R}),\\
&-x(1\mathrm{R},2\mathrm{W})-x(2\mathrm{R},2\mathrm{W})+x(2\mathrm{W},3\mathrm{R})+x(2\mathrm{W},4\mathrm{R})=0 \quad (\text{node } 2\mathrm{W}),\\
&-x(1\mathrm{W},3\mathrm{R})-x(2\mathrm{W},3\mathrm{R})+x(3\mathrm{R},3\mathrm{W})=0 \quad (\text{node } 3\mathrm{R}),\\
&-x(1\mathrm{R},3\mathrm{W})-x(2\mathrm{R},3\mathrm{W})-x(3\mathrm{R},3\mathrm{W})+x(3\mathrm{W},4\mathrm{R})=0 \quad (\text{node } 3\mathrm{W}),\\
&-x(1\mathrm{W},4\mathrm{R})-x(2\mathrm{W},4\mathrm{R})-x(3\mathrm{W},4\mathrm{R})=-1 \quad (\text{node } 4\mathrm{R}).
\end{aligned}
\tag{3.17}
$$

Using the substitutions described above to simplify (3.4), we obtain the following vector equations for the twelve flow variables in terms of the twelve arc variables:

$$
\begin{bmatrix} u_1\\ s_1\\ w_1\\ I_1\\ u_2\\ s_2\\ w_2\\ I_2\\ u_3\\ s_3\\ w_3\\ I_3 \end{bmatrix}
=0.\ x(1\mathrm{R},1\mathrm{W})+
\begin{bmatrix} 0\\ d_1\\ d_1\\ 0\\ d_1\\ 0\\ \cdot\\ \cdot\\ \cdot\\ \cdot\\ \cdot\\ 0 \end{bmatrix}
x(1\mathrm{R},2\mathrm{W})+
\begin{bmatrix} 0\\ d_1\\ d_1\\ 0\\ 0\\ d_1+d_2\\ d_2\\ 0\\ d_1+d_2\\ 0\\ 0\\ 0 \end{bmatrix}
x(1\mathrm{R},3\mathrm{W})
$$

$$+\begin{bmatrix} d_1 \\ 0 \\ \mathrm{d}_1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} x(1\mathrm{W}, 2\mathrm{R}) + \begin{bmatrix} d_1+d_2 \\ 0 \\ \mathrm{d}_1 \\ \mathrm{d}_2 \\ 0 \\ 0 \\ \mathrm{d}_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} x(1\mathrm{W}, 3\mathrm{R})$$

$$+\begin{bmatrix} d_1+d_2+d_3 \\ 0 \\ d_1 \\ d_2+d_3 \\ 0 \\ 0 \\ d_2 \\ d_3 \\ 0 \\ 0 \\ d_3 \\ 0 \end{bmatrix} x(1\mathrm{W}, 4\mathrm{R}) + \text{etc.} \tag{3.18}$$

The equations for the twelve fixed charge variables in terms of arc variables are quite similar, and begin in this manner:

$$\begin{bmatrix} z(u_1) \\ z(s_1) \\ z(w_1) \\ z(I_1) \\ z(u_2) \\ z(s_2) \\ z(w_2) \\ z(I_2) \\ z(u_3) \\ z(s_3) \\ z(w_3) \\ z(I_3) \end{bmatrix} = 0.\ x(1\mathrm{R}, 1\mathrm{W}) + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} x(1\mathrm{R}, 2\mathrm{W}) + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} x(1\mathrm{R}, 3\mathrm{W})$$

$$+\begin{bmatrix}1\\0\\1\\0\\ \cdot\\ \cdot\\ \cdot\\ \cdot\\ \cdot\\ \cdot\\ \cdot\\ 0\end{bmatrix} x(1\mathrm{W}, 2\mathrm{R}) + \text{etc.} \tag{3.19}$$

To write (3.19), we assumed positive demands (all $d_i > 0$) and place a unit wherever a symbolic nonzero quantity occurred in the corresponding vector in (3.18), with one exception: A production fixed charge is assessed only on an arc from a "white" to "red" node, following (3.15) and (3.16). Moreover, a zero quantity in (3.18) always gives a corresponding zero in (3.19).

When some of the demands d_i are zero, they can be replaced by zero, and (3.18) simplifies. Then the same rules apply to obtain expressions for fixed charges.

As (3.17)-(3.19) shows, even for $r = 3$ periods, the substitutions, which give the original flow and fixed charge variables in terms of the arc variables, are complex. The equations (3.18) and (3.19) need not be actually part of the constraints, but used only to substitute out for original variables in any remaining constraints which involve original flow variables only (i.e., the remaining constraints cannot involve original fixed charge variables).

Example 3.6. In our second example, we shall augment the simple lot-sizing model by a different form of backlogging. We consider a setting in which delaying a customer order due in period i by two periods, until period $(i+2)$, is significantly more costly than delaying a customer order from period i to $(i+1)$ plus delaying the customer order from period $(i+1)$ to $(i+2)$, either in terms of per unit charges or fixed charges, or possibly both charges. Furthermore, it is not possible in this scenario to delay a customer order by more than two periods. This scenario is designed for cases where customers become increasingly impatient with additional delays, beyond simply additive effects, to the point of possibly cancelling their orders.

It is important to realize that there is more than one product flow network for this problem.

One such network is given in Fig. 2. In terms of extreme points, the individual trees involved are more complex than those for Fig. 1, due to the possibility of *non-consecutiveness* in the time periods supplied out of manufacture in a given period. For example, in an extreme point solution it is possible for period 3 to supply

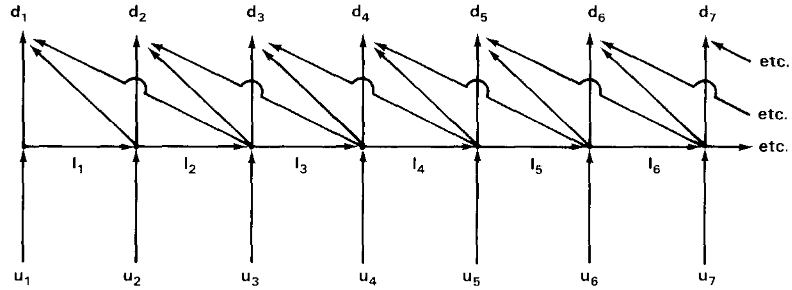

Fig. 2.

demand for periods 1, 3, 4 and 6; with period 2 supplied by period 1 manufacture; and period 7 supplying demand for period 5 as well as for itself and possibly other periods.

A second, and different, product flow network matching exactly the same informal verbal description is shown in Fig. 3. It retains all the flow options as in Fig. 2, but since product backlogged two periods from period j to $(j-2)$ now passes through the demand arc in period $(j-1)$, additional flows are now possible. Specifically, product made in period $(j-2)$ can be inventoried into period $(j-1)$, and then used as backlog to satisfy demand in period $(j-2)$, should that ever be worthwhile. Of course, the backlogging arc, from the intersection with the arc in period $(j-1)$ to the demand point in period $(j-2)$, will have on it very high costs in an actual application, since that arc is intended for use by twice-backlogged demand. Thus such a convoluted flow would not be worthwhile. We also assume a cost structure in which once-backlogged demand from period $(j-1)$ to $(j-2)$ is cheaper than the second backlog cost, of twice backlogging from period j to $(j-2)$.

Interestingly enough, the extreme points of the network with more options (Fig. 3) are easier to describe and involve consecutivity, which will allow us a more economical sharp MIP formulation. For instance, due to the additional point of intersection in period 2 (indicated by an arrow in Fig. 3), if period 3 supplies period 1 via backlogging, it must also supply period 2 in an extreme point solution. Indeed, if

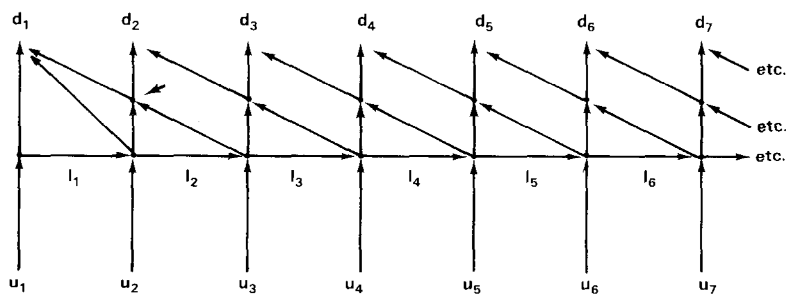

Fig. 3.

flow from either manufacture in periods 1 or 2 were to supply demand in period 2, these flows would cross the new intersection point (unless they went on to period 3, which is not possible in a tree). By doing so, the property that only one positive source is possible in a tree (Lemma 3.4) would be violated, so this possibility is ruled out.

In terms of viewing the extreme points as built up from consecutive decisions, with each decision corresponding to an arc in some network and the sequence of decisions corresponding to a path, the path network of Example 3.1 can be used with the stipulation that $i \leq j+2$. Thus fewer arcs occur than before, but again $2r+1$ node conservation equations. We leave the detailed implementation as an exercise, together with its extension to nonadditive backlogging costs for any number of periods.

This completes our two examples for the specific principles developed in this section. Many extensions of these ideas are possible, and we mention one.

In place of a single polytope (3.8), a finite union of such polytopes can be treated similarly, again using Theorem 3.1 to obtain compact sharp formulations. Such unions arise, for example, when logical conditions are imposed on the variables of a polytope.

4. Results on sums of unions

In this section, we provide results regarding programs of the form:

$$\begin{aligned} &\min \quad cx+dy, \\ &\text{s.t.} \quad Ax \geq b-Du, \quad x \in S = \sum_{k=1}^{t} S_k, \ x \in R^p, \end{aligned} \tag{4.1}$$

where, for each k,

$$S_k = \bigcup_{h \in I_k} W_{kh}, \tag{4.2}$$

and each W_{kh} is a polyhedron. Our purpose is to generalize some well-known results to a broader representability setting (as e.g. in [8]).

Let the number of rows of A be m. As regards the representation of S_k, we first consider the case that it is given by

$$\begin{aligned} x^{(k)} \in S_k \quad \Leftrightarrow \quad & x^{(k)} = \textstyle\sum_h x^{(kh)}, \ 1 = \sum_h \lambda_{kh}, \text{ all } \lambda_{kh} \in \{0,1\}, \\ & A^{(kh)}x^{(kh)} + B^{(kh)}y^{(kh)} \geq b^{(kh)}\lambda_{kh} \text{ for } h \in I_k, \end{aligned} \tag{4.3)$_k$}$$

where in $(4.3)_k$ any polyhedral representation of W_{kh} may be used. By Corollary 3.4 for $k_1 = k_2 = 1$, $(4.3)_k$ is a sharp representation of S_k.

In (4.1), the constraints "$Ax \geq b-Du$" are intended to represent the case of "soft" constraints "$Ax \geq b$", in which the qth constraint can be relaxed at a cost of d_q per unit of relaxation. However, the constraints "$x \in S$" are not to be relaxed.

An instance of (4.1) occurs in the *shop loading problem* in which job k can be done on machine i in time σ_{ki}, and each job is to be uniquely assigned to exactly one machine. Here S_k represents the alternative assignments of job k, so that $I_k = \{1, \dots, p\}$ for all k (recall $x = (x_1, \dots, x_p)$) and $W_{ki} = \{\sigma_{ki} e_i\}$ is a singleton set, where e_i denotes the ith unit vector. In this context, x_i will represent the total amount of work (in terms of time required) assigned to machine i, not allowing for set-ups or precedence requirements, etc., if any. If we have u consists of a single variable u_i, with $d_1 = 1$, and the matrix D consists of the single vector $e^{\mathrm{T}} = (1, 1, \dots, 1)$ with all entries unity, $A = -I$, and $b = 0$, then (4.1) is the problem of minimizing the makespan (i.e., latest completion time) of completing all jobs. Indeed, with these settings $Ax \geq b - Du$ is equivalent to $x_i \leq u_1$ for all i, so (4.1) is the problem of minimizing $\max_i x_i$.

Several generalizations of the shop loading problem are accomodated by (4.10). These include: (1) Charges for machine utilization (i.e., nonzero c_i); (2) time utilization on machine i is significant only after a given period (nonzero b_i); (3) there are multiple alternatives for distributing job k among the machines, and these do not necessarily involve "proportionality factors" of the machines (more general sets W_{kh}).

Lemma 4.1. *Suppose that there is a constant* $K \geq 0$ *such that, for all* k *and all* $\bar{x}^{(k)} \in \operatorname{conv}(S_k)$, $x^{(k)} \in S_k$, *there is* $u^{(k)}$ *with*

$$A(x^{(k)} - \bar{x}^{(k)}) \geq -Du^{(k)}, \qquad du^{(k)} + c(x^{(k)} - \bar{x}^{(k)}) \leq K. \tag{4.4}$$

Let the linear relaxation of (4.1) *be solved* (*i.e.*, "$\lambda_{kh} \in \{0, 1\}$" *is replaced by* "$\lambda_{kh} \geq 0$" *in* $(4.3)_k$) *for an extreme point solution, obtaining quantities* $\bar{x}$, $\bar{x}^{(k)}$, $\bar{x}^{(kh)}$, $\bar{y}^{(kh)}$, $\bar{\lambda}_{kh}$. *Let* $x^{(k)}$ *be chosen from among those* $w^{(kh)} = \bar{x}^{(kh)}/\bar{\lambda}_{kh}$ *with* $\bar{\lambda}_{kh} > 0$ (*but otherwise arbitrarily*).

Then for $x = \sum_k x^{(k)}$ *we have* $x \in S$, *and* u *can be chosen so that* $cx + du$ *does not exceed the value of the linear relaxation by more than* $K \min\{m, p\}$.

Proof. First, note that, since $w^{(kh)} \in W_{kh}$ we have $x^{(k)} \in S_k$. Thus $x \in S$ is clear.

We claim that not more than $t + p$ of the quantities $\bar{\lambda}_{kh}$ are positive.

Once this claim is proven; then for at least $t - p$ of the t sets S_k we have exactly one index $h^* \in I_k$ with $\bar{\lambda}_{kh^*} > 0$ (hence $\bar{\lambda}_{kh^*} = 1$), so that $x^{(k)} = \bar{x}^{(k)} = \bar{x}^{(kh^*)}$. Denote by M the index set of these sets, and by E the index set of the remaining at most p "exceptional" sets. We have (since $x - \bar{x} = \sum_{k \in E} (x^{(k)} - \bar{x}^{(k)})$):

$$\begin{aligned} Ax &= A\bar{x} + A(x - \bar{x}) \\ &= A\bar{x} + \sum_{k \in E} A(x^{(k)} - \bar{x}^{(k)}) \\ &\geq b - D\bar{u} - \sum_{k \in E} Du^{(k)}. \end{aligned} \tag{4.5}$$

In (4.5), we used (4.4) and the facts that $\bar{x}^{(k)} \in \operatorname{conv}(S_k)$, $x^{(k)} \in S_k$. Similarly,

$$cx + d\Big(\bar{u} + \sum_{k \in E} u^{(k)}\Big) = c\bar{x} + d\bar{u} + \sum_{k \in E} (du^{(k)} + c(x^{(k)} - \bar{x}^{(k)}))$$
$$\leq c\bar{x} + d\bar{u} + pK \quad (\text{as } |E| \leq p).$$

We now prove the claim.

Put $I_k^+ = \{h \in I_k \mid \bar{\lambda}_{kh} > 0\}$ and recall that $w^{(kh)} = \bar{x}^{(kh)}/\bar{\lambda}_{kh}$ for $h \in I_k^+$. The linear system

$$\bar{x} = \sum_{k=1}^{t} \Big(\sum_{h \in I_k^+} \lambda_{kh} w^{(kh)}\Big) + \sum_{k=1}^{t} \Big(\sum_{h \in I_k^+} \bar{x}^{(kh)}\Big), \tag{4.6}$$
$$1 = \sum_{h \in I_k^+} \lambda_{kh}, \quad \text{for all } k,$$
$$\lambda_{kh} \geq 0, \quad \text{for all } h \in I_k, \text{ all } k$$

in $p+t$ constraints is solved by the $\bar{\lambda}_{kh}$ (see (4.1), (4.2), $(4.3)_k$). If more than $p+t$ of the $\bar{\lambda}_{kh}$ were positive, there would exist a nontrivial set of multipliers θ_{kh}, such that both $(\bar{\lambda}_{kh} \pm \theta_{kh},\ h \in I_k,\ k = 1, \ldots, t)$ solve (4.6). Both these modifications extend to entire solutions to (4.1), (4.2), and $(4.3)_k$ by putting $x^{(kh)} = \bar{x}^{(kh)}$ for $h \in I_k^+$, and by setting, for $h \in I_k^+$, $x^{(kh)} = (\bar{\lambda}_{kh} \pm \theta_{kh}) w^{(kh)}$, $y^{(kh)} = (\bar{\lambda}_{kh} \pm \theta_{kh}) \bar{y}^{(kh)}/\bar{\lambda}_{kh}$, $x^{(k)} = \sum_h x^{(kh)}$. However, these two distinct solutions have the given solution as their average, contradicting the extremality of the latter. This contradiction proves the claim.

We also claim that not more than $t+m$ of the quantities $\bar{\lambda}_{kh}$ are positive. Just as the previous claim implies a value not exceeding that of the linear relaxation by more than pK, this claim implies an excess of not more than mK. The two claims, therefore, imply the desired result.

The latter claim is established by considering the linear system

$$A\Big(\sum_{k=1}^{t} \Big(\sum_{h \in I_t^+} \lambda_{kh} w^{(kh)}\Big)\Big) \geq b - D\bar{u}, \tag{4.7}$$
$$1 = \sum_{h \in I_k^+} \lambda_{kh}, \quad \text{for all } k,$$
$$\lambda_{kh} \geq 0, \quad \text{for all } h \in I_k, \text{ all } k$$

in $m+t$ constraints. We leave the details to the reader. □

Theorem 4.2. *Suppose that these hypotheses hold:*

(1) *The diameter of S_k is bounded by $\Delta \geq 0$, independently of k;*
(2) *for any vector v, there exists u with $Du \geq v$.*

Then there is a constant K (depending only on Δ, A, D, c and d) such that the difference in value between the program (4.1), (4.2), $(4.3)_k$ *and that of its linear relaxation does not exceed $K \min\{m, p\}$, and moreover the procedure of Lemma* 3.1 *finds a solution to this program with value within this bound of the optimum.*

Proof. By hypothesis (1), if $x^{(k)} \in S_k$ and $\bar{x}^{(k)} \in \text{conv}(S_k)$, no coordinate of $x^{(k)} - \bar{x}^{(k)}$ has absolute value exceeding 2Δ. From this, a bound can be found on that of any coordinate of $A(x^{(k)} - \bar{x}^{(k)})$, independent of k. From hypothesis (2), a vector u can be found with $A(x^{(k)} - \bar{x}^{(k)}) \geq -Du$, independent of k. It is then easy to verify the existence of K such that (4.4) holds. This result then follows from Lemma 3.1. □

In the shop loading example, hypothesis (2) of Theorem 4.2 is immediate. Hypothesis (1) requires that all $\sigma_{ki} \leq \Delta$ for some $\Delta \geq 0$, i.e., that jobs do not arrive with time needs on the machine which is least efficient for them, which increases without any bound. In this context, if also there exists $\delta > 0$ with $\delta \leq \sigma_{ki}$ for all k, i, then the gap result of Theorem 4.2 also yields an approximation result. Indeed, since the optimal value of the linear relaxation is at least $t\delta/p$, the fraction error is $pK \min\{mp\}/t\delta \to 0$ as $t \to +\infty$, i.e., it is $O(1/t)$. See [24] for an analysis of shop loading with a probability distribution on job requirements, where the fraction error is shown to be stochastically $O(1/t^2)$.

References

[1] K.J. Arrow and F.H. Hahn, General Competitive Analysis (Holden-Day, San Francisco, CA, 1971).

[2] E. Balas, Disjunctive programming: Cutting-planes from logical conditions, in: O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds., Nonlinear Programming 2 (Academic Press, New York, 1975) 279–312.

[3] E. Balas, Disjunctive programming, in: P.L. Hammer, E.L. Johnson and B.H. Korte, eds., Discrete Optimization II (North-Holland, Amsterdam, 1979) 3–52.

[4] E. Balas, Disjunctive programming and a hierarchy of relaxations for discrete optimization problems, SIAM J. Algebraic Discrete Methods 6 (1985) 466–486.

[5] I. Barany, T.J. van Roy and L.A. Wolsey, Strong formulations for multi-item capacitated lot sizing, Management Sci. 30 (1984) 1255–1261.

[6] I. Barany, T.J. van Roy and L.A. Wolsey, Uncapacitated lot sizing: The convex hull of solutions, Math. Programming Stud. 22 (1984) 32–43.

[7] I. Barany, T.J. van Roy and L.A. Wolsey, Valid linear inequalities for fixed charge problems, Oper. Res. 33 (1985) 842–861.

[8] J.F. Benders and J.A.E. van Nunen, A property of assignment type mixed integer linear programming problems, Oper. Res. Lett. 2 (1983) 47–52.

[9] C.E. Blair, Two rules for deducing valid inequalities for zero-one programs, SIAM J. Appl. Math. 31 (1977) 614–617.

[10] C.E. Blair and R.G. Jeroslow, A converse for disjunctive constraints, J. Optim. Theory Appl. 25 (1978) 195–206.

[11] J.W.S. Cassels, Measures of the nonconvexity of sets and the Shapley–Folkman–Starr theorem, Math. Proc. Cambridge Philos. Soc. 78 (1975) 433–436.

[12] U.C. Choe, Arc-path approaches to fixed-charge network problems, Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, GA (1983).

[13] A. Claus, A new formulation for the travelling salesman problem, SIAM J. Algebraic Discrete Methods 5 (1984) 21–25.

[14] G.D. Eppen and R.K. Martin, Solving multi-item capacitated lot sizing problems using variable redefinition, Graduate School of Business, University of Chicago (1985).

[15] L.R. Ford Jr and D.R. Fulkerson, Flows in Networks (Princeton, Univ. Press, Princeton, NJ, 1962).
[16] T. Ibaraki, Integer programming formulation of combinatorial optimization problems, Discrete Math. 16 (1976) 39-52.
[17] R.G. Jeroslow, Cutting-plane theory: Disjunctive methods, in: Annals of Discrete Mathematics 1 (North-Holland, Amsterdam, 1977) 293-330.
[18] R.G. Jeroslow, Representability in mixed integer programming I: Characterization results, Discrete Appl. Math. 17 (1987) 223-243.
[19] R.G. Jeroslow, Two mixed integer programming formulations arising in manufacturing management (1987).
[20] R.G. Jeroslow, Alternative formulations of mixed-integer programs, Ann. Oper. Res. 12 (1988) 241-276.
[21] R.G. Jeroslow, Logic-based decision support: Mixed integer model formulation, Ann. Discrete Appl. Math. 40 (1989).
[22] R.G. Jeroslow and J.K. Lowe, Modelling with integer variables, Math. Programming Stud. 22 (1984) 167-184.
[23] R.G. Jeroslow and J.K. Lowe, Experimental results on the new techniques for integer programming formulations, J. Oper. Res. Soc. 36 (1985) 393-403.
[24] J.K. Lenstra, A.H.G. Rinnooy Kan and L. Stougie, A framework for the probabilistic analysis of hierarchical planning systems, Stichting Mathematisch Centrum, Amsterdam (1983).
[25] M.Y. Leung, Polyhedral structure of capacitated fixed charge problems and a problem in delivery route planning, Ph.D. Dissertation, MIT, 1985.
[26] J.K. Lowe, Modelling with integer variables, Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA (1984).
[27] R.K. Martin, Generating alternative mixed-0/1 linear programming models using variable redefinition, Graduate School of Business, University of Chicago, Chicago, IL (1984).
[28] R.K. Martin, A sharp representation for certain index sets, Private communication (1986).
[29] R.K. Martin, A sharp polynomial size linear programming formulation of the minimum spanning tree problem, Graduate School of Business, University of Chicago, Chicago, IL.
[30] R.R. Meyer, On the existence of optimal solutions to integer and mixed-integer programming problems, Math. Programming 7 (1974) 223-235.
[31] R.R. Meyer, Integer and mixed-integer programming models: General properties, J. Optim. Theory Appl. 16 (1974) 191-206.
[32] R.R. Meyer, Mixed-integer minimization models for piecewise-linear functions of a single variable, Discrete Math. 16 (1976) 163-171.
[33] R.R. Meyer, A theoretical and computational comparison of "equivalent" mixed integer formulations, Naval Res. Logist. Quart. 28 (1981) 115-131.
[34] R.R. Meyer, M.V. Thakkar and W.P. Hallman, Rational mixed integer and polyhedral union minimization models, Math. Oper. Res. 5 (1980) 135-146.
[35] G. Owen, Cutting planes for programs with disjunctive constraints, J. Optim. Theory Appl. 11 (1973) 49-55.
[36] M.W. Padberg, T.J. van Roy and L.A. Wolsey, Valid linear inequalities for fixed charge problems, Oper. Res. 33 (1985) 842-861.
[37] Y. Pochet and L.A. Wolsey, Lot-size models with backlogging: Strong reformulations and cutting planes, CORE Discussion Paper 8618, Université Catholique de Louvain, Louvain.
[38] R.L. Rardin and V. Choe, Tighter relaxations of fixed charge network flow problems (1979).
[39] R.T. Rockafellar, Convex Analysis (Princeton Univ. Press, Princeton, NJ, 1970).
[40] T.J. van Roy and L.A. Wolsey, Valid inequalities and separation for uncapacitated fixed charge networks, Oper. Res. Lett. 4 (1985) 105-112.
[41] A. Schrijver, On cutting planes, in: Annals of Discrete Mathematics 9 (North-Holland, Amsterdam, 1980) 291-296.

[42] J. Stoer and C. Witzgall, Convexity and Optimization in Finite Dimensions I (Springer, Berlin, 1970).
[43] A.F. Veinott Jr, Minimum concave-cost solution of Leontief substitution models of multi-facility inventory systems, Oper. Res. 17 (1969) 262–291.
[44] H.M. Wagner, A postscript to "Dynamic problems of the firm", Naval Res. Logist. Quart. 7 (1960) 7–12.
[45] H.M. Wagner and T.M. Whitin, Dynamic version of the economic lot size model, Management Sci. 5 (1968) 89–96.
[46] H.P. Williams, Model Building in Mathematical Programming (Wiley/Interscience, New York, 2nd ed., 1985).
[47] R.T. Wong, Integer programming formulations of the travelling salesman problem, in: Proceedings of 1980 IEEE International Conference on Circuits and Computers (1980) 149–152.
[48] W.I. Zangwill, A backlogging model and a multi-echelon model of a dynamic economic lot size production system: A network approach, Management Sci. 15 (1969) 506–527.
[49] W.I. Zangwill, A deterministic multi-period production scheduling model with backlogging, Management Sci. (to appear).